```r
#Required packages
#Most packages are already installed for any social scientist running ecological
inference
install.packages(c("tidyverse","tidygeocoder","eiCompare","wru","tigris","data.ta
ble","readxl"))

#Libraries
library(tidyverse)
library(tidygeocoder)
library(eiCompare)
library(wru)
library(tigris)
library(data.table)
library(readxl)

#Load in voter file excel spreadsheet
#Sample R script points to November 2022 vote history file produced by Galveston
County
data <- read_xlsx("~/DEFS00031066-DEFS00031099/DEFS00031066-DEFS00031099
_orig/DEFS011/NATIVES/0001/DEFS00031067.xlsx")

#Geocode using tidygeocoder
data <- geocode(
  data,
  address = "Residence Address",
  method = "geocodio")

#In the event a voter address is incomplete or missing
no_address <- data %>% filter(is.na(lat))
data <- data %>% filter(!is.na(lat))

#Load up Texas census blocks to join to voter file
tx_blocks <- blocks("TX", year = 2020)

#n = Number of voters
n <- length(data$`Voter Name`)

#Create ID for voter file to shapefile merge
data$unique_id <- 1:n

#Merge voter and Texas census blocks
data <- merge_voter_file_to_shape(data,
                                  tx_blocks,
                                  coords = c("long","lat"),
                                  voter_id = "unique_id")

#Revert back to dataframe object from shapefile for WRU
data <- as.data.frame(data)

#If any voters had missing address merge back in
data <- full_join(data, no_address, by = NULL)

#Create two-character abbreviation for state for WRU
data$state <- "TX"

#Rename geofips columns for WRU
data <- data %>%
  rename(county = COUNTYFP20,
```

1

Def_Add_Ex3

```
            tract = TRACTCE20,
            block = BLOCKCE20)

#If separate surname column does not already exist in voter file, separate "Voter
Name" column into surname and first / middle name columns
data <- separate(data, col = `Voter Name`, into = c("surname", "first_middle"),
sep = ",")

#Probabilistic race/ethnicity estimates from WRU
data <- predict_race(
  voter.file = data,
  surname.only = TRUE,
  census.geo = "block",
  year = "2020")

#Aggregate to precinct-level to feed into eiCompare
data <- precinct_agg_combine(
  data,
  group_col = "Precinct")
```

2